AD_____

Award Number: W81XWH-12-1-0050

TITLE: Development of Noninvasive Biomarkers for Diagnosing and Monitoring Nonindolent Prostate Cancer

PRINCIPAL INVESTIGATOR: Daniella Bianchi-Frias, Ph.D.

CONTRACTING ORGANIZATION: Fred Hutchinson Cancer Research Center
Seattle, WA 98109-1024

REPORT DATE: April 2013

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
April 2013	Annual	15 March 2012 – 14 March 2013

4. TITLE AND SUBTITLE

Development of Noninvasive Biomarkers for Diagnosing and Monitoring Nonindolent Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-12-1-0050

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Daniella Bianchi-Frias, Ph.D.

E-Mail: amoreno@fhcrc.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Fred Hutchinson Cancer Research Center
Seattle, WA 98109-1024

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
While prostate cancer (PCa) is the most common solid organ malignancy in men, only 20–30% progress to metastatic disease. Men with indolent disease are offered a treatment plan that involves active surveillance (AS). However, there is a significant risk of under-grading the tumor and repeated invasive biopsies are required. We hypothesized that transcripts associated with high Gleason grade cancers are quantifiable in urine samples from men with prostate cancer, and that measurements of grade-associated transcripts will reflect the presence of higher-grade non-indolent tumors. By gene expression analysis (from microdissected Gleason-pattern (GP) 3 and GP4 PCa), in combination with publically available Gleason-associated transcriptional profiles, we have created a 46-gene panel that differentiates high Gleason from low Gleason grade PCa. Moreover, we have found that up-regulation of several GP-associated transcripts, such as RELN, associate with adverse clinical outcomes. We validated the GP4-associated upregulation of candidate genes by qPCR. Additionally, we have started to measure by qPCR the transcript levels for 6-genes in urine sediments from patients undergoing biopsy. Although, a significant difference exists between negative biopsy and PCa for two genes, no significant differences were found between GS6 and GS≥8 biopsies. Urine sediments from patients undergoing radical prostatectomy are necessary to test their accuracy in predicting high versus low grade tumors. The discovery of high grade-associated transcripts in urine from patients with presumed indolent prostate cancer could substantially improve accurate staging of prostate tumors and clinical management decisions.

15. SUBJECT TERMS
Prostate Cancer, Gleason, Active Surveillance, Urine, Biomarkers

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	21	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

INTRODUCTION

Prostate cancer (PCa) is a common disease with incidence rates that rise dramatically with advancing age. Though clearly neoplastic, the vast majority of prostate cancers behave in an indolent fashion. Men with indolent disease are currently offered a treatment plan that involves deferred intervention or active surveillance (AS). However, two of the major limitations and concerns for AS strategies involve under-sampling of existing tumor foci (which represents a significant risk of under-grading the tumor) and requirements for invasive (biopsy) methods for cancer assessments. Thus, to this end, there is a significant clinical need for the development of biomarkers that can be measured noninvasively and can distinguish between men undergoing AS that develop high grade prostate cancer and men with indolent disease.

The Gleason scoring system is considered one of the most powerful prognosticators in PCa[1, 2], thus this proposal will test the hypothesis that transcripts associated with high Gleason grade cancers are quantifiable in urine samples from men with prostate cancer, and that measurements of grade-associated transcripts will reflect the presence of higher-grade non-indolent tumors. We expect that a urine-based assay of GP-associated transcripts will identify occult higher grade cancers that either were missed on initial diagnostic biopsies or that emerged/evolved over time (biological progression).

BODY

Task 1. Define cohorts of transcript alterations that associate with high grade (Gleason pattern 4-5) versus low grade (Gleason pattern 3) cancers

A number of studies have reported gene expression signatures correlating with Gleason grade using expression arrays[3-8]. To date however, consensus among Gleason Pattern (GP)-associated transcriptional profiles in localized prostate cancer has not been determined, in part because technology at that time did not facilitate genome wide analyses. The heterogeneity of the prostate samples assessed on each study (various representations of tumor cells and ratios of tumor cells-to normal glands-and-stroma, versus microdissected samples highly enriched with tumor cells) may also contribute to the lack of consistency among studies.

Using older, partial-genome microarrays (PEDB cDNA microarray, 3708-unique-genes) and LCM samples, our group has previously identified a 86-gene classifier capable of distinguishing low-grade (Gleason Pattern 3: GP3) from high-grade (Gleason Pattern 4 & 5: GP4 and GP5, respectively) cancers[7]. To expand our analysis, and create a more comprehensive GP-associated gene panel, I undertook an additional independent discovery effort to define GP-associated transcripts using contemporary full-genome expression arrays (Agilent 44K oligonucleotide microarray, 19643-unique-genes) and profiled transcripts across a separate set of microdissected prostatic tissue (Figure1).

Twenty five samples were from non-neoplastic prostate epithelium adjacent to tumor, referred herein as benign samples. The tumor samples included GP3 cells (n=15) from: 9 GS (3+3) cases, 4 GS(3+4) and 1 GS(4+3) cases, and GP4 cells (n=13) from: 4 GS(3+4) cases, 3 GS(4+3) cases, and 6 GS(4+4) cases. These Gleason grades correspond to the scores assigned to the tissue blocks from which the cells were microdissected and in some cases differed from the Clinical Gleason score assigned to the radical prostatectomy tissue (RP-Gleason). Patient demographic characteristics are show in Table 1. Of the 25 cases, I excluded one case due to poor microarray hybridization. Pathological review of the LCM images verified the intended GP3 and GP4 cells collected, respectively.

For the Agilent microarray experiment, probe labeling and hybridization was performed following the manufacturer's suggested protocols and fluorescent array images were collected using the Agilent DNA microarray scanner G2565BA. Data was loess normalized within arrays and quantile normalized between arrays in R using the Limma Bioconductor package. Our new Agilent microarray data consisted of two-channel ratios of the benign, GP3 and GP4 microdissected prostate tissue, all hybridized against a common reference sample.

Unsupervised cluster analysis using the top 1000 most variable genes, clearly grouped the samples into two branches: branch I represented by benign samples and branch II represented by cancer samples, regardless of Gleason grade (Figure 2A). As expected, prostate cancer associated transcripts, such as AMACR and HPN, were significantly up regulated in cancer compared to benign samples (Figure 2B). This result confirms, at the molecular level, that an accurate microdissection of the intended cell type was achieved.

To explore the relationship between GP3 and GP4 samples, we performed Principal Component Analysis (PCA) for all the genes in the arrays (Figure 2C). PCA clearly grouped a subset of genes that discriminated benign and cancer samples, confirming that the major differences resulted from the differential expression of large numbers of genes between the benign and cancer samples and not by Gleason grade, as observed in the dendogram described above. Nevertheless, within the Gleason samples, PCA could partially separate GP3 from GP4 samples as shown in Figure 2C (arrowheads). To further characterize the relationships between GP3 and GP4 samples, the interquartile range of virtual head-to-head ratios of each cancer sample (to the patient-matched normal) was computed and the top 1000 most variable genes were clustered using Pearson correlation distance and average linkage (Figure 3). Cancer samples were grouped into 4 major branches: Branch I is represented by GP3 samples microdissected from RP-Gleason 3+4 and no biochemical recurrence (BRC), branch II and III represented by GP4 samples from RP-Gleason 4+4 with biochemical recurrence and metastatic outcomes, and branch IV and V represented by an intermediate group of GP3 and GP4 with RP-Gleason 3+4 and 4+3 samples and some recurrence cases. These observations suggest that a molecular signature can distinguish low-grade, low risk PCa (branch I) from the most aggressive high-grade, high risk PCa (branches II and III). However, the histological defined Gleason specific transcripts do not represent a dichotomous variant, and that the expression is rather a continuum from less aggressive to more aggressive cancers as represented by branch IV.

To identify genes whose expression in GP4 significantly differed from GP3 samples we used the Statistical Analysis of Microarray (SAM) program[9] and applied an unpaired, two-sample t-tests analysis and controlled for multiple testing by estimation of q-values using the false discovery rate (FDR) method. This analysis defined a cohort of 620 mRNAs with

GP-associated differential expression (Figure 4). For the identification of candidate-urine-biomarkers, I focused on transcripts highly expressed in GP4 rather than down-regulated in GP4. We believe this approach (Up in GP4) will facilitate the quantitative analysis in urine samples, since a low or lack of expression of a gene does not rule out low representation of prostate cancer cells in urine, or a unsuccessful qPCR reaction. Still, the inclusion of few (one, or two) down-regulated genes in GP4 among several up-regulated genes could also be valuable when considering a gene-panel. Among the significantly differentially expressed genes, RGS5 was the most upregulated gene with a 15 fold-enrichment in GP4 compared to GP3 and Normal (Table 2). Further, within the significantly up-regulated gene list, I have found that several GP-associated transcripts, such as RGS5 RELN and C5orf30 or a combination of them associate with adverse clinical outcomes, such as biochemical recurrence following primary therapy, as expected based on the known adverse outcomes associated with higher Gleason scores[10] (Figure 5).

In order to compare our new full-genome expression array with our partial genome array, we merged the Agilent and PEDB Gleason datasets. After spot quality assessment, the merged data contained 3011 unique genes in common between both platforms. The True et al. PEDB microarray data consisted of two-channel head-to-head ratios of laser-capture microdissected Gleason 3, 4, and 5 patterns of cancer against patient-matched benign epithelium. Using these ratios we compared GP3 with either GP4 alone or GP4 combined with GP5. Our new Agilent microarray data consisted of two-channel ratios of laser-capture microdissected epithelium and GP3 and GP4 cancer cells, all hybridized against a common reference sample. A low but significant correlation coefficient of 0.23 ($p<0.0001$) between the two distinct microarray experiments was determined using the scored T-test.

Initially we compared the original ratios for each platform to identify common differentially expressed genes between GP3 and GP4. Additionally, we created virtual head-to-head ratios of each cancer sample compared to the patient-matched normal and compared the groups again. Overlap of genes with q-values less than 10% were computed and shown by Venn diagrams (Figure 6). Seventy genes were significantly up-regulated in both studies and only 6 genes down-regulated (Fig 6A and 6B) using the original ratios calculated for each platform. MAOA, whose higher expression was previously confirmed at the protein level, was among the genes in common between both platforms. Other genes significantly up-regulated in high Gleason in the PEDB data set, such as DAD1, were not up-regulated in GP4 in our new Agilent array. Using the virtual head-to head-ratios to create the Venn Diagrams, the overlap between both studies was significantly reduced to only 13 genes up-regulated and 3 down-regulated in GP4 compared to GP3. (Figure 6C). This low overlap, besides being affected by the different platforms used and the low number of common genes (3011 genes), suggest that a GP-associated signature is not a robust phenotype, even though histological defined Gleason patter cells were laser captured microdissected in both studies.

In order to generate a comprehensive GP-associated gene candidate list, we integrated our two array datasets described above, with the meta-analysis-determined grade-associated transcripts and selected those mRNAs consistently up-regulated in GP4 PC relative to benign epithelium and GP3 PC (Table 2). The meta-analysis consisted of a cross-study normalized matrix of mRNA expression comprising data from 251 benign prostate tissue samples, 852 primary prostate cancers samples, and 47 metastatic samples. With this matrix, we created a gene list of transcripts differentially expressed

between GS6 and GS8-9. The GP-associated candidates were generated based on: i) most significant difference between low- and high-grade cancer in Agilent dataset; ii) highest overexpression in GP4 iii) previously validated; vi) overlap with PEDB dataset and meta-analysis; and v) Preferentially express in prostate tissue compared to bladder and kidney tissues (evaluated in tissue-specific portals: BioGPS and TiGER). This effort produced a GP-associated cohort of 46 transcripts that I have started to evaluate for their potential as urine biomarkers. Additionally, I will include any emerging targets that are reported during this next period.

Task 2. Develop specific assays to quantitate Grade-associated transcripts in tissue and in urine samples

Validation of a Gleason pattern associated transcript panel in prostate tissue. For the purpose of refining the GP-associated biomarker panel, I have begun the development of qPCR assays for the quantitative determination of transcript levels in tissue and urine. I have started with the 46-marker panel described above and have constructed 33 assays to date. Aliquots of the same samples that were amplified and labeled to generate the Agilent microarray results (Cohort 2, C2) were also analyzed by qPCR. Twenty-five of thirty-three genes tested confirmed the microarray results. Representative results for the qPCR analyses are shown in Figure 7 and p-values for all the genes testes are in Table 2. To validate the differential expression in an independent cohort, aliquots of the same samples that were amplified and labeled to generate the original PEDB expression profile were used for qPCR analyses (Cohort 1, C1.). Eleven of the Twenty-three genes tested to date, were significantly up-regulated in GP4/GP5 compared to benign and GP3 PC, validating the results obtained in cohort 2 (C2) (Table 2). Representative results for 4 markers are shown in Figure 8C. I will evaluate the expression of candidate genes in a second independent cohort consisting of RNAs extracted from 20 frozen section containing >70% cancer with Gleason 3+3 (n=20) and twenty of Gleason 4+4, respectively.

Validation of a Gleason pattern associated transcript panel in urine sediments. I have also begun the development of qPCR for the quantitative determination of transcript levels in urine from patients presenting for needle biopsy. As an initial experiment in urine samples, we tested 5 candidate genes in a small cohort of urine sediments from biopsy cases. Within this cohort, n=5 cases were GS6; n=5 GS ≥8 and n=5 had negative biopsy. The transcript levels for these 5 genes were readably detectable in the urine sediment by qPCR, demonstrating the feasibility of our assay using SYBG qPCR. The cycle number for PSA ranged between 27-32 Ct. In order to confirm the presence of prostate cells in urine and normalize the cycle number obtained by qPCR, we used the prostate specific marker: PSA (KLK3). Different studies have reported several normalization strategies in which, only PSA is used to normalize the Ct, or a house keeping gene (e.g. GAPDH) is used to normalize for total RNA in combination with PSA which will internally normalize for prostate cells $(Ct_{PSA} + Ct_{GAPDH})/2 - Ct_{Variable}$). After employing these two normalization strategies, none of the candidate genes tested were statistically significantly up-regulated in bx GS≥8, compared to either NEG or GS6 (see Table 3 for p-values for each gene tested and Figure 8 for representative results). Nevertheless, the box plots in Figure 8E demonstrate a trend for higher expression of the candidate genes in GS≥8. When an unpaired t-test analysis is performed between negative versus positive biopsy, the expression of HOXD3 and WNK3 were significantly up-regulated in positive biopsy samples, regardless of its Gleason score.

The lack of significant alterations of the candidate genes between NEG, GS6 and GS≥8 from urine sediments could have been influenced by multiple reasons: 1) High variability in expression levels between samples, as shown in Figure 8A to 8D the GP4 samples have a wide range of expression per case (red points), thus with o low number of samples significance cannot be reached; 2) Very low volumes of high-grade cancer that may not release quantities of cells and resultant transcripts sufficient for detection 3) Expression of genes in urothelial cells; 4) The use of a single prostate specific marker PSA, which is highly variable among samples (Ct range 27-32) could affect the final normalized results, since the normalized Ct are highly influenced by the PSA concentration, I suggest including few more prostate specific transcripts that could be used for normalization. In order to identify those candidates, I will use our gene expression data sets and identified genes that were not significantly altered between benign and GP3 ad GP4 such and that are not expressed in bladder, kidney and immune cells.

Another imperative aspect when developing the urine assay is to use the most appropriate samples to develop the assay and create a model. The low certainty on the accuracy between the biopsy and clinical Gleason scores, underscores the need of using urine sediments from radical prostatectomy cases (from which the clinical Gleason score is assigned) in order to develop an accurate model of urine biomarkers for high grade prostate cancer detection.

Since GP-associated transcript levels represent a continuum of expression with higher levels correlating with high Gleason grade, and that do not behave as a dichotomous variable, a multivariate logistic regression analysis might prove to be valuable to define significance among several genes. Thus, we expect that a panel of grade-associated markers will be required.

If we do not find correlation between the GP markers and significant cancers on biopsy or prostatectomy, we will combine data from the Gleason marker assays with urinary TMPRSS2:ERG and PCA3 data[11] and establish multivariate models that may perform better at distinguishing apparently indolent disease.

KEY RESEARCH ACCOMPLISHMENTS:

- I have identified a 46-gene expression profile that correlates with high Gleason grade prostate cancer (Task 1)
- 25/33 genes tested by qPCR confirmed the Agilent microarray results from prostatectomy tissue specimens (Task 2)
- 11/23 genes tested by qPCR to date, validated the GP-associated expression using an independent prostatectomy tissue specimens cohort (Task 2)
- A subset of the 46-gene candidates identified associate with prostate cancer recurrence (Task 2)
- Established an RNA-based urine assay by qPCR using urine sediments from biopsies (Task 2)

REPORTABLE OUTCOMES:

- Presentation: "Detecting high grade-specific transcripts in urine to improve active surveillance". Prostate Cancer Meeting, FHCRC, Seattle, WA.

- Gleason-grade associated gene expression database

CONCLUSION:

I have identified a molecular signature that underlies the histological classification of prostate cancer Gleason grades using both, gene expression analyses and cross-comparison between publically available datasets. I have identified gene outliers within the GP4 group, that could have the potential to discriminate low versus high Gleason grade when use as a gene-panel. Further, I have found that several GP-associated transcripts correlate with adverse clinical outcomes.

Although, cluster analysis revealed that a molecular signature can distinguish low-grade, low risk PCa from the most aggressive high-grade, high risk PCa, --the GP-expression phenotype is not a robust, nor a dichotomous variant, and that gene expression levels are rather a continuum from less aggressive to more aggressive cancers. Thus, taken together, this data confirms the concept of implementing a biomarker-panel rather than a single biomarker for the assessment of non-indolent PCa in urine.

In order to test GP-associated candidate genes in urine samples, it is essential to perform the assays using urine collected from patients undergoing radical prostatectomies, rather than biopsy, in order to be confident of the Gleason score assigned and thus incorporate that information into the model.

REFERENCES:

1. Epstein JI: An update of the Gleason grading system, J Urol 2010, 183:433-440
2. Gleason DF, Mellinger GT: Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging, J Urol 1974, 111:58-64
3. Agell L, Hernandez S, Nonell L, Lorenzo M, Puigdecanet E, de Muga S, Juanpere N, Bermudo R, Fernandez PL, Lorente JA, Serrano S, Lloreta J: A 12-gene expression signature is associated with aggressive histological in prostate cancer: SEC14L1 and TCEB1 genes are potential markers of progression, Am J Pathol 2012, 181:1585-1594
4. Liu P, Ramachandran S, Ali Seyed M, Scharer CD, Laycock N, Dalton WB, Williams H, Karanam S, Datta MW, Jaye DL, Moreno CS: Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells, Cancer Res 2006, 66:4011-4019
5. Ross AE, Marchionni L, Vuica-Ross M, Cheadle C, Fan J, Berman DM, Schaeffer EM: Gene expression pathways of high grade localized prostate cancer, Prostate 2011,
6. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, Kalyana-Sundaram S, Wei JT, Rubin MA, Pienta KJ, Shah RB, Chinnaiyan AM: Integrative molecular concept modeling of prostate cancer progression, Nat Genet 2007, 39:41-51
7. True L, Coleman I, Hawley S, Huang CY, Gifford D, Coleman R, Beer TM, Gelmann E, Datta M, Mostaghel E, Knudsen B, Lange P, Vessella R, Lin D, Hood L, Nelson PS: A molecular correlate to the Gleason grading system for prostate adenocarcinoma, Proc Natl Acad Sci U S A 2006, 103:10991-10996
8. Vanaja DK, Cheville JC, Iturria SJ, Young CY: Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression, Cancer Res 2003, 63:3877-3882
9. Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response, Proc Natl Acad Sci U S A 2001, 98:5116-5121
10. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, Antipin Y, Mitsiades N, Landers T, Dolgalev I, Major JE, Wilson M, Socci ND, Lash AE, Heguy A, Eastham JA, Scher HI, Reuter VE, Scardino PT, Sander C, Sawyers CL, Gerald WL: Integrative genomic profiling of human prostate cancer, Cancer Cell 2010, 18:11-22

11. Tomlins SA, Aubin SM, Siddiqui J, Lonigro RJ, Sefton-Miller L, Miick S, Williamsen S, Hodge P, Meinke J, Blase A, Penabella Y, Day JR, Varambally R, Han B, Wood D, Wang L, Sanda MG, Rubin MA, Rhodes DR, Hollenbeck B, Sakamoto K, Silberstein JL, Fradet Y, Amberson JB, Meyers S, Palanisamy N, Rittenhouse H, Wei JT, Groskopf J, Chinnaiyan AM: Urine TMPRSS2:ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA, Sci Transl Med 2011, 3:94ra72

APPENDICES:

SUPPORTING DATA:

Table 1. Patient demographics*	
Gleason score, clinical (N)	
<6	3
7	18
>8	3
Gleason score, sample block (N)	
<6	10
7	8
>8	6
Gleason pattern captured (N)	
3	13
4	12
Matched Benign (N)	24
PSA at RP (ng/ml)	6.9 (2.5–63.4)
Pathological Stage (N)	
T2cN0	12
T3aN0	4
T3xN+	6
Tumor Volume	3 (0.6-10)
Biochemical Recurrence	
No	16
YES	4
Nadir	3

* Where applicable, Median(Range) is listed

Table 2. Gleason-associated transcripts: 46-gene panel

Gene Symbol	Agilent microarray (Cohort 2, C2)		PEDB microarray (Cohort 1, C1)		qRT-PCR (p –values)			Prognostic Value** (p<0.05*) Taylor et al.	Meta-analysis fold increase in GS9 vs. GS6
	q-value (%)	GP4-Fold increase (G3/N vs. G4/N)	q-value (%)	GP4-Fold increase (G3/N vs. G4/N-G5/N)	C2. (G4/N vs. G3/N)	C1. (G5/N vs. G3/N)	C1. (G4/N vs. G3/N)		
RGS5*	0	14.3	n/a	n/a	0.001	0.009	0.150	YES	n/a
GRIN3A	0	6.8	n/a	n/a	0.001	0.842	0.910	no	1.4
FRY	0	5.8	n/a	n/a	0.000	0.105	0.256	no	1.1
IL1RAPL1	0	5.7	1	1.3	0.013	0.968	0.601	no	1.0
NRP1	0	5.5	n/a	n/a	n/a	n/a	n/a	no	1.1
CXCR7*	0	5.3	14	1.3	0.000	0.002	0.079	no	1.0
SSTR1	0	5.2	n/a	n/a	0.067	n/a	n/a	YES	1.0
HOXD3	0	5.0	n/a	n/a	0.000	0.224	0.231	YES	1.0
LRRN1*	0	4.7	n/a	n/a	0.000	0.035	0.089	no	1.6
RFX6	1	4.4	n/a	n/a	0.039	n/a	n/a	YES	1.2
FCGR3A	0	4.1	10	2.0	n/a	n/a	n/a	no	1.1
GRIK1	4	4.1	6	1.4	n/a	n/a	n/a	YES	-1.0
C5orf30*	0	3.8	n/a	n/a	0.001	0.048	0.075	YES	1.1
MCTP1	0	3.8	n/a	n/a	0.002	n/a	n/a	no	1.0
MID1	0	3.8	n/a	n/a	0.144	n/a	n/a	no	1.0
PECAM1	0	3.3	n/a	n/a	n/a	n/a	n/a	no	1.1
ONECUT2	1	3.2	n/a	n/a	0.085	n/a	n/a	YES	1.0
HEG1*	0	3.2	n/a	n/a	0.000	0.044	0.143	no	-1.0
CXCL12	2	3.2	48	1.1	n/a	n/a	n/a	no	-1.1
WFDC5	11	2.9	n/a	n/a	0.186	n/a	n/a	YES	-1.1
HIGD1B	0	2.8	n/a	n/a	n/a	n/a	n/a	no	-1.0
C11orf80	0	2.8	n/a	n/a	0.001	n/a	n/a	no	1.1
RELN	7	2.8	n/a	n/a	0.066	n/a	n/a	YES	1.1
UTS2D	2	2.7	n/a	n/a	0.005	n/a	n/a	YES	1.0
ZMIZ1*	0	2.7	n/a	n/a	0.000	0.015	0.041	no	1.1
CILP	3	2.6	n/a	n/a	n/a	n/a	n/a	YES	-1.0
PDZD2	0	2.5	n/a	n/a	0.004	0.145	0.355	no	1.0
WNK3*	0	2.5	n/a	n/a	0.004	0.004	0.011	no	1.2
RAB23*	0	2.5	n/a	n/a	0.001	0.002	0.035	no	-1.0
KCTD12*	0	2.3	3	2.1	0.000	0.001	0.038	no	1.1
IMPA1	1	2.3	9	1.4	n/a	n/a	n/a	no	1.1
CDON	0	2.3	n/a	n/a	n/a	n/a	n/a	no	1.0
BICC1	0	2.2	n/a	n/a	0.100	0.082	0.427	no	-1.0
FOLH1*	1	2.2	36	1.5	0.001	0.010	0.012	YES	1.7

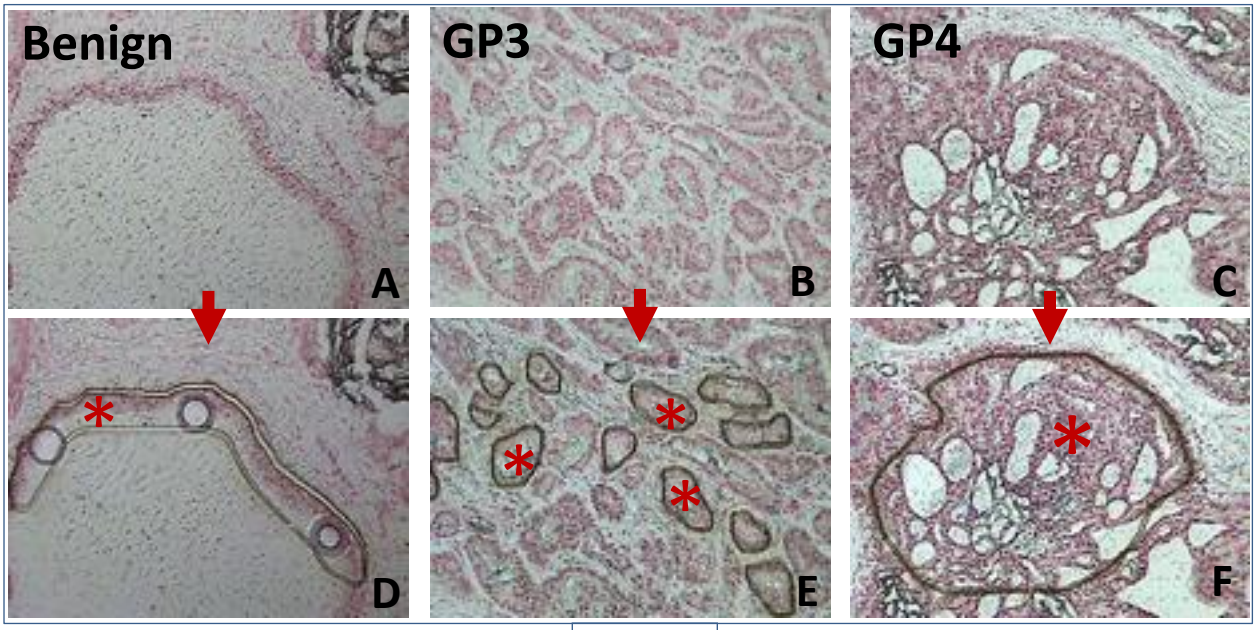
Continuation..Table 2									
CLDN8	0	2.2	1	1.6	n/a	n/a	n/a	no	1.1
TJP1	0	2.1	35	-1.2	0.001	0.178	0.420	no	1.1
CPEB4*	0	2.1	n/a	n/a	0.002	0.026	0.044	no	-1.0
MAOA	8	2.0	0	2.3	n/a	n/a	0.04	no	1.1
NCOA1	0	1.9	n/a	n/a	0.000	0.076	0.092	YES	1.0
UTRN	0	1.8	n/a	n/a	0.001	0.236	0.147	no	1.1
HOXC6	6	1.7	n/a	n/a	0.439	0.937	0.568	no	1.4
PPFIA2	32	1.7	46	-1.3	0.042	n/a	n/a	no	1.1
STMN1	53	1.0	23	1.4	0.320	n/a	n/a	no	1.1
ZNF492	14	-1.0	n/a	n/a	n/a	n/a	n/a	no	-1.0
CLEC14A	45	-1.1	n/a	n/a	0.003	n/a	n/a	YES	1.1
PSGR2	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1.5

*Genes validated by qPCR ; **n/a:** not present, not measured

** Prognostic value as determined using Taylor et al[10] data set in the cBioPortal for Cancer Genomics site

Table 3. p-values of qPCR analysis from urine sediments

Gene Symbol	(CtPSA+CtGAPDH)/2 - Ct Variable p-value				CtPSA- CtVariable p-value			
	*NEG vs GS6	NEG vs GS9	NEG vs Ca	GS9 vs GS6	NEG vs GS6	NEG vs GS9	NEG vs Ca	GS9 vs GS6
WNK3	**0.0281**	0.0999	**0.0441**	0.6683	0.177	0.184	0.144	0.736
HOxd3	0.0651	0.0989	**0.0175**	0.9551	0.119	0.151	**0.045**	0.949
RELN-F2	0.1129	0.1269	0.0610	0.5864	0.269	0.209	0.161	0.676
RGS5	0.1188	0.1016	0.0537	0.5051	0.286	0.185	0.159	0.627
GRIN3A	0.1876	0.1495	0.1131	0.4594	0.456	0.296	0.297	0.602
ZMIZ1	0.5617	0.5166	0.4548	0.9296	0.638	0.574	0.544	0.936
GAPDH	na	na	na	na	0.900	0.943	0.906	0.955



Sample processing

- Samples: Benign, n=24; GP3, n=14; and GP4 n=13)
- RNA extraction, amplification and labeling
- Hybridization into Agilent 44K arrays

Gene Expression Analysis

- two-channel ratios: benign, GP3 or GP4 hybridized against a common reference.
- Data loess normalized within arrays and quantile normalized between arrays
- Statistical analysis of gene expression: SAM program (unpaired, two-sample t-test controlled for multiple testing

Candidate selection

- Q-value, Fold change method
- Highest IQR
- Overlap with PEDB arrays
- Overlap with meta-analysis
- Literature

Figure 1. Flow chart demonstrating experimental design. Pre (A,B and C) and post (D, E, F) - captures images, asterisk mark microdissected areas within the prostate tissue.

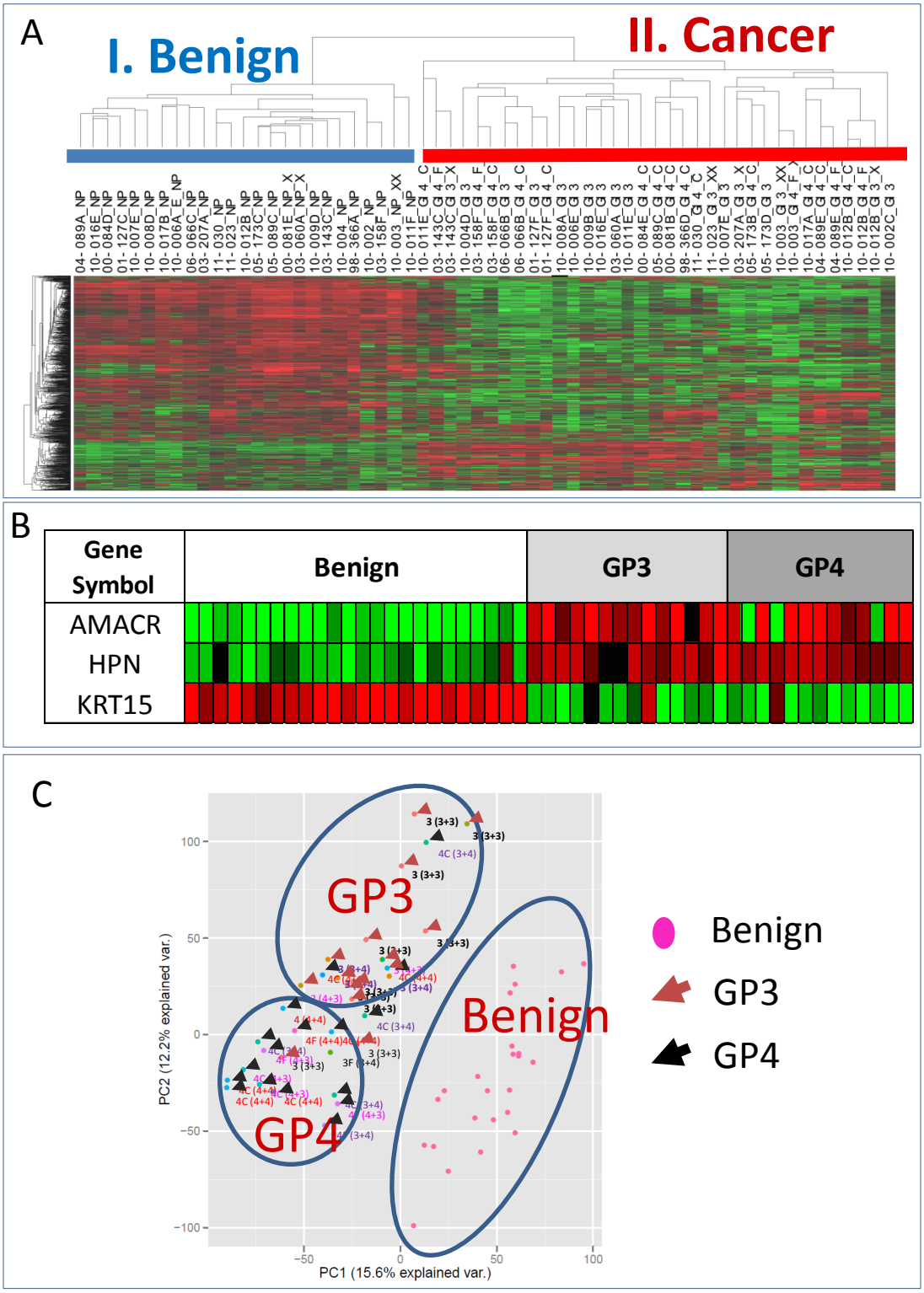


Figure 2. Hierarchical cluster analysis , Heatmap and Principal Component Analysis (PCA) of prostate samples. (A) Hierarchical cluster analysis of benign (n=24), GP3 (n=14), and GP4 (n=13) samples. Mean-centered gene expression ratios are shown by a log2 color scale (Red represent up- and green down- regulated genes compare to median values). (B) HeatMap for prostate cancer-associated genes across all samples. (C) PCA analysis across all samples.

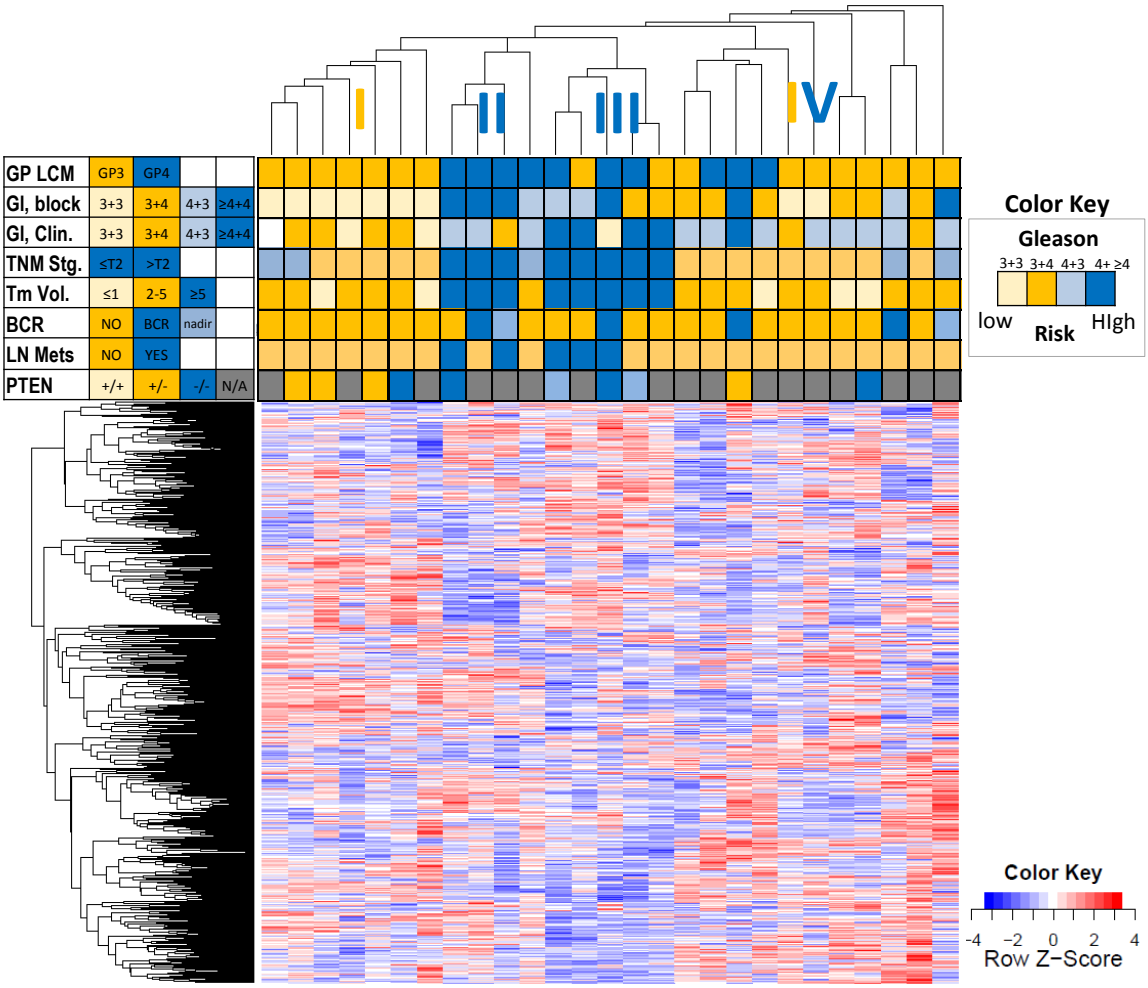


Figure 3. (A) Hierarchical cluster analysis of prostate cancer Gleason samples. Clinicopathological features associated with individual tumor samples are indicated by yellow and blue boxes below the dendrogram (grey indicate missing data). **GP LCM** indicates Gleason pattern microdissected; **GI, block**: Gleason score in tissue block; **GI, Clin**.: Clinical Gleason score; **TNM Stg**.: pathological stage. **Tm Vol**: Tumor volume; **BCR**: Biochemical Recurrence (PSA rise after surgery). **LN Mets**: positive lymph nodes or clinical metastasis. **PTEN**: Genomic deletion of PTEN locus (+/- Heterozygous, -/- Homozygous deletion). Blue indicates, high grade, high risk advanced stage PCa and yellow indicates low grade, low risk PCa. HeatMap, log2 rations for the 1000 most variable genes. Red up-regulated and Blue: down-regulated genes expression.

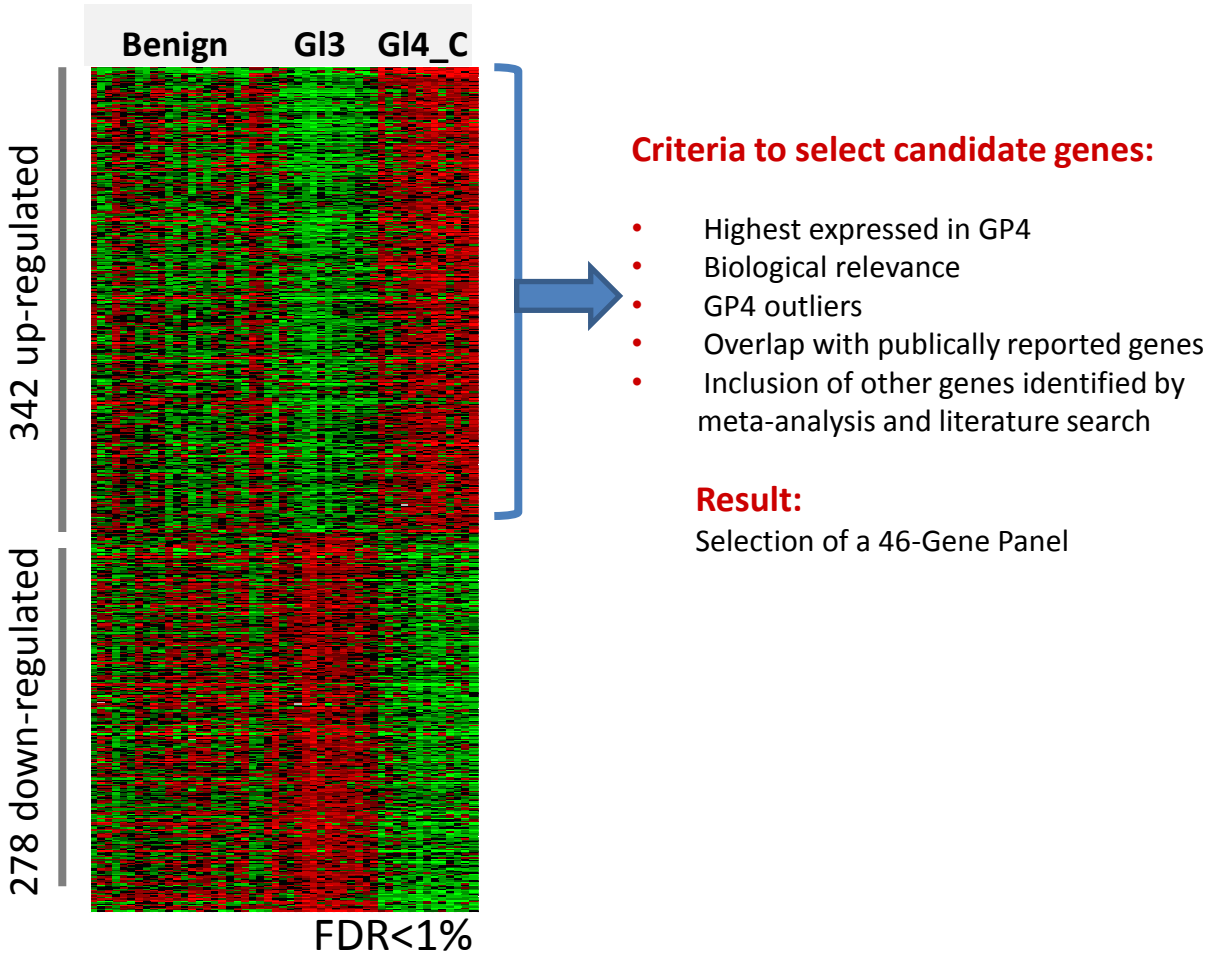


Figure 4. Prostate Cancer Gleason Pattern (GP)-associated gene expression. Heatmap of transcript abundance level differences determined by full-genome microarray analysis, across microdissected benign epithelium, GP3
and GP4 prostate cancer

Prostate Adenocarcinoma (MSKCC, Cancer Cell 2010)/Primary Tumors with mRNA: (131) /User-defined List/2 genes :EXP>2)

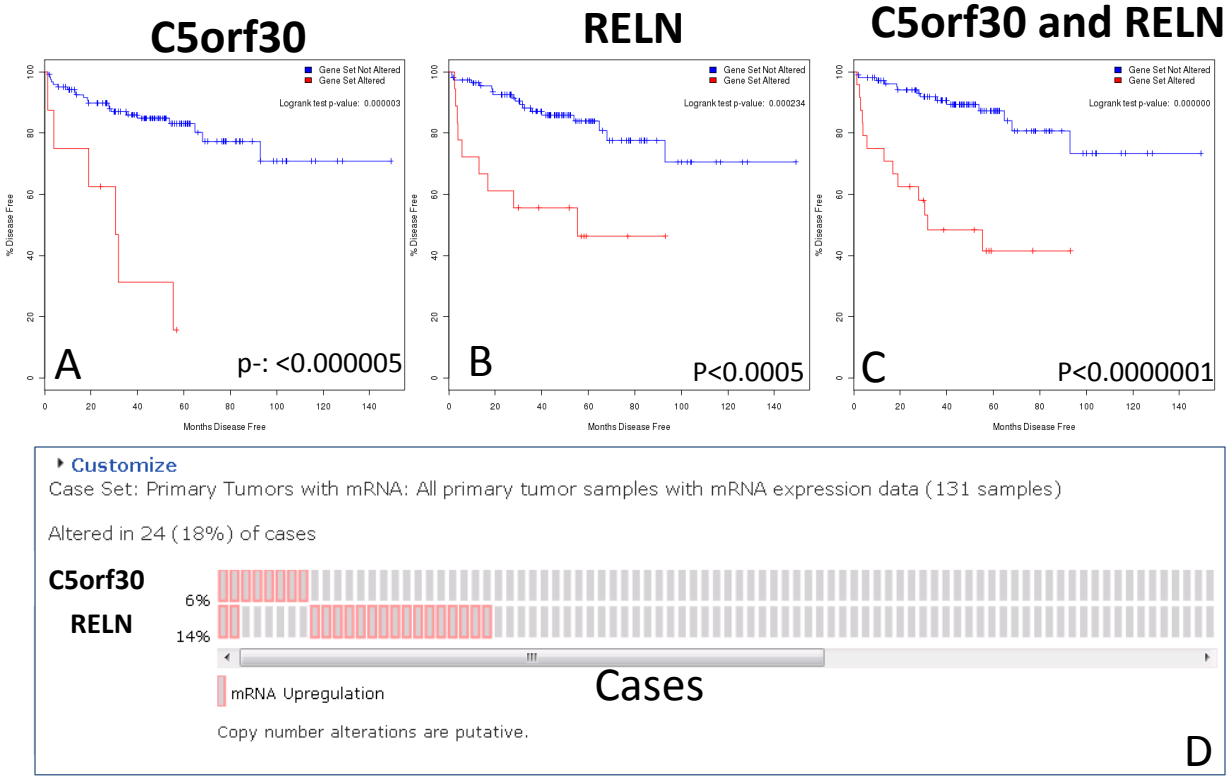


Figure 5. Kaplan–Meier survival analysis of Taylor *et al*. cohort, assessing correlation of GP-associated overexpression of gene candidates with survival outcomes. (A-B) Overexpression (>2 z-scores) of C5orf30 (A) and RELN (B) can segregate patients into good (blue) and poor (red) prognostic categories. (C) A 2-gene-panel model is better able to prognosticate recurrence. (D) Cases in Taylor et al. cohort, that overexpress the candidate genes. Grey bars represent independent prostate cancer samples. Note the lack of overlap between the overexpressed gene C5orf30 and RELN, among all prostate cancer cases, favoring the concept for the use of a gene-panel to asses a wide range of tumor, potentially revealing non-indolent prostate adenocarcinomas subtypes.

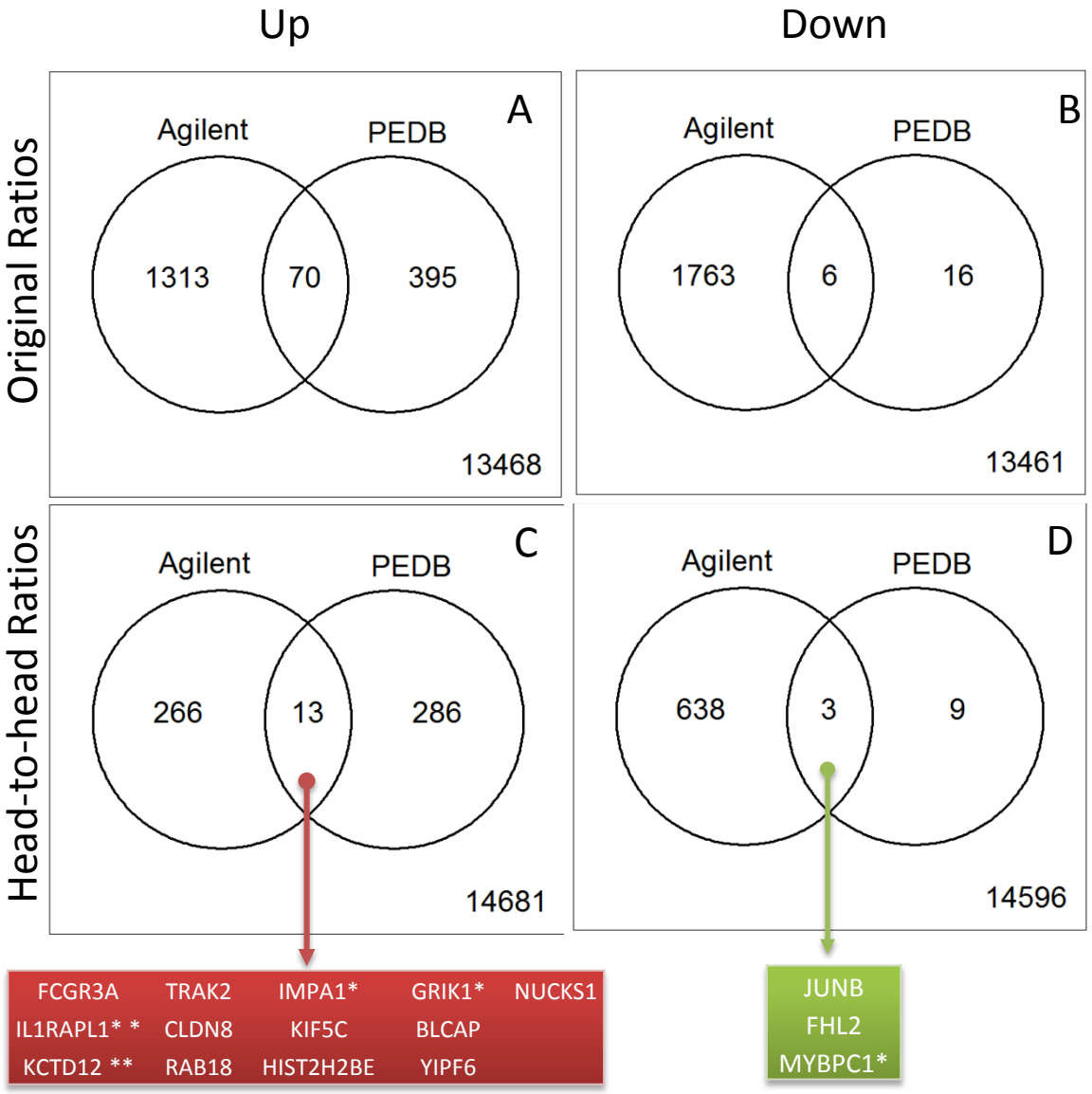


Figure 6. Overlap of genes with q-values less than 10% between Agilent and PEDB Gleason-associated transcriptional profiles. (A, B) up- and down- regulated genes, respectively defined by t-test scores using original Log2 Ratio. (C,D) up- and down- regulated genes, respectively defined by t-test scores using virtual head-to-head ratio in Agilent dataset.

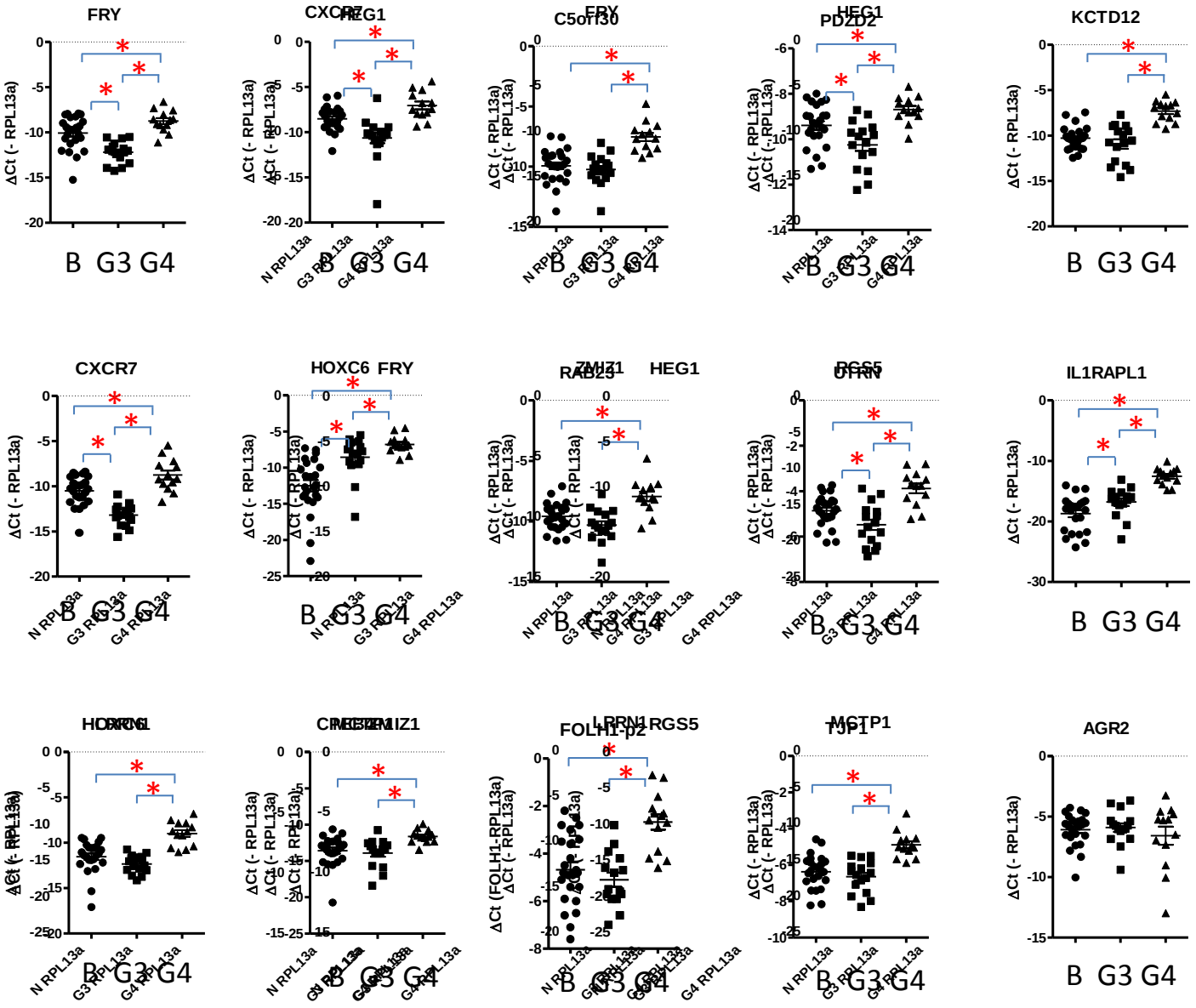
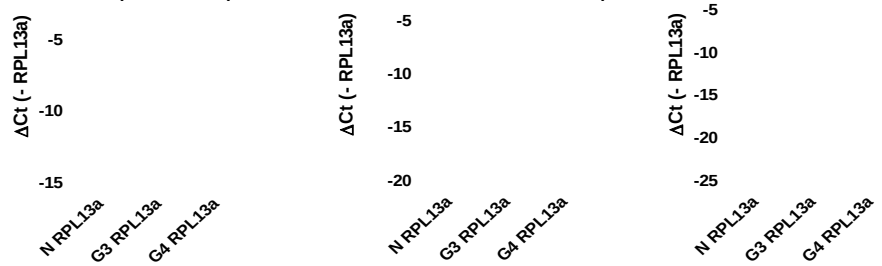


Figure 7. Confirmation of GP-associated transcripts. qPCR assays were developed to confirm GP-associated transcripts. Shown are 15 representative genes. Each data point represents an independent PC sample. * is p<0.05 for the indicated comparison.



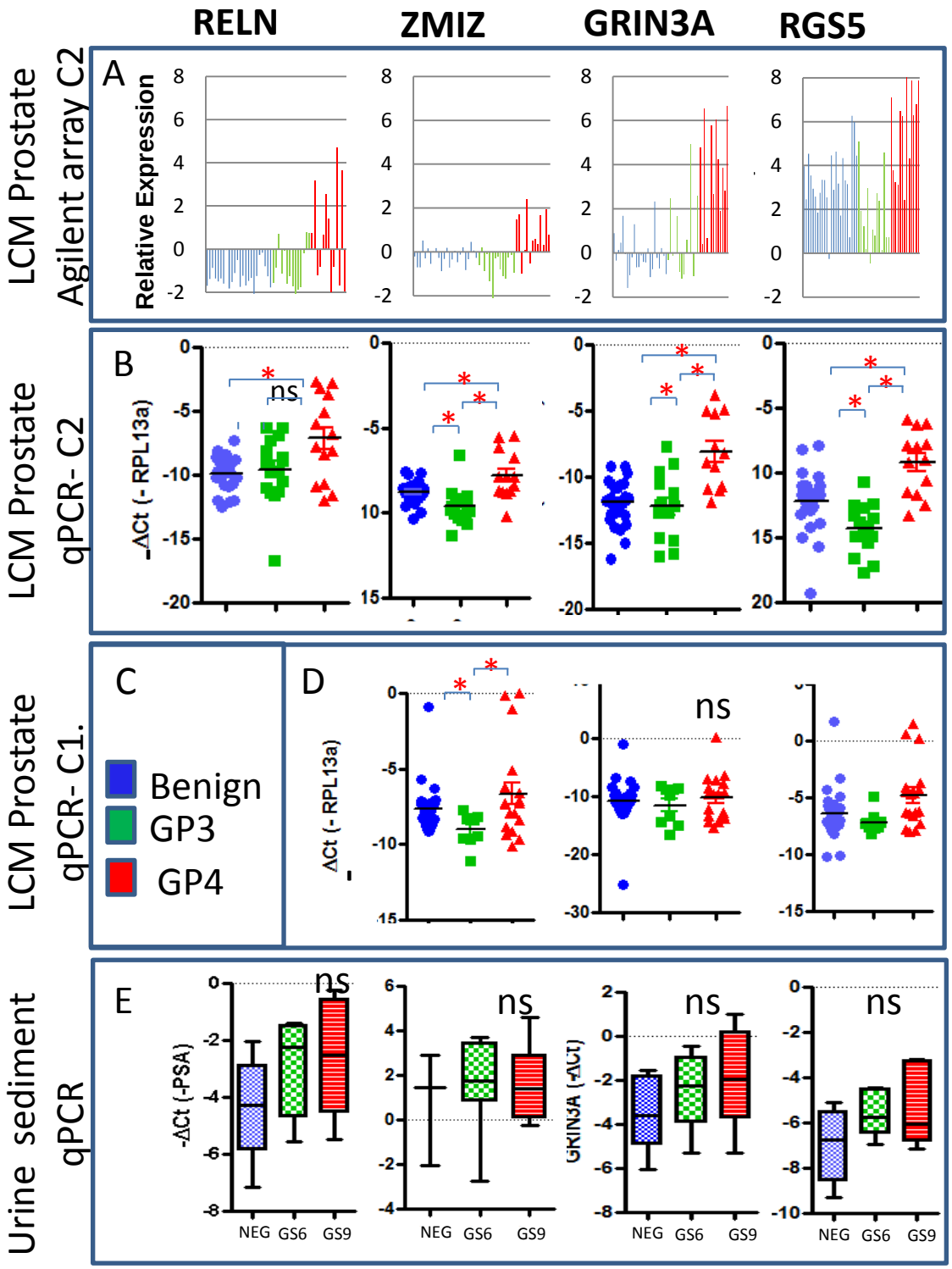


Figure 8. Characterization of candidate genes as GP-associated urine biomarkers. (A) Relative Log2Ratio levels on Agilent arrays, to a common gold standard reference. (B-D) qPCR assays on cDNA from prostate microdissected tissue from 2 independent cohorts. C1: Cohort 1, samples used in PEDB. C2: Cohort 2, samples used for Agilent arrays. Expression in benign (blue), LCM Gleason 3 (green) and LCM Gleason 4 (red). (E) qPCR was performed on cDNA from urine sediments, obtained from patients presenting for needle biopsy. Biomarker expression in patients with negative needle biopsies (blue), or patients with prostate cancer GS6 (green) and GS9 (red). Normalization was performed using delta Ct, with candidate gene normalized to urine PSA expression. ns: non-significant unpaired t-test analysis $p>0.05$. * is $p<0.05$ for the indicated comparison. Shown are 4 representative genes.